

# Information Centric Networks

Daniel Sousa <201605245> Tiago Carvalho <201605259> Tiago Coelho <201604170>

**Resumo**—Este artigo tem como finalidade expor um novo paradigma de endereçamento de conteúdo na Internet conhecido como Information Centric Networks.

Vamos começar por descrever a operação desta tecnologia a um nível alto, explicar a sua importância na Internet de hoje em dia, e para concluir, partilharemos algumas *frameworks*, que consideramos relevantes, a implementar este inovador princípio.

**Index Terms**—ICN; P2P; conteúdo; escalabilidade

## I. INTRODUÇÃO

A Internet que hoje em dia conhecemos foi arquitetada nos anos 60 e 70, com o intuito de resolver o problema de partilha de recursos [1] (sobretudo de correio eletrónico e ficheiros) em contexto académico [2], algo que intrinsecamente se traduzia em baixos volumes de tráfego. O modelo adotado na altura, e que ainda domina a Internet para partilha de recursos, é o Cliente-Servidor, onde um número arbitrário de nós de origem (os clientes) estabelecem uma conexão (normalmente, TCP) com um único nó destino (o servidor), de forma a descarregar algum tipo de recurso. Um leitor atento consegue aperceber-se de uma falha inerente, relativamente óbvia, deste modelo: sempre que o servidor ficar indisponível (*offline*), o recurso que pretendemos descarregar, e potencialmente muitos outros hospedados na mesma máquina, vão ficar também estes indisponíveis.

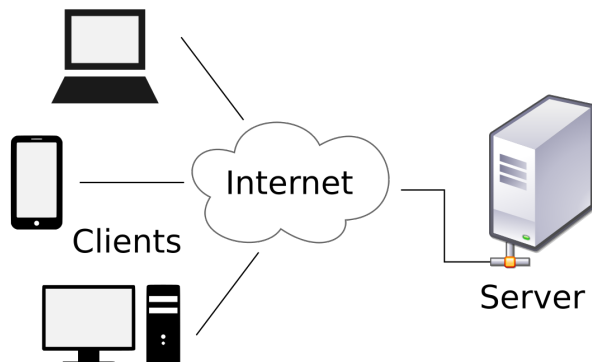


Figura 1: Modelo Cliente-Servidor [3]

Para além disso, um cliente deverá ter conhecimento prévio do caminho padrão para o recurso que pretende descarregar (e.g. pedidos HTTP GET [4]), que poderá sofrer alterações ao longo do tempo, apesar do conteúdo para o qual este aponta ser de natureza estática (e.g. ficheiros de multimédia), e com este tipo de arquitetura, pode ser difícil

garantir a integridade da origem do recurso que um dado cliente pretende descarregar, que por vezes será um fator determinante da sua respetiva validação (e.g. extratos de *e-banking*).

Em suma, vimos que existem problemas estruturais na Internet “tradicional” relacionados com [1]:

- **Disponibilidade** – A entrega rápida e contínua de certas classes de conteúdo (e.g. multimédia) exige mecanismos específicos ao domínio da aplicação como redes P2P<sup>1</sup> ou CDNs<sup>2</sup>.
- **Independência de localização** – É necessário haver uma correspondência entre nó destino e caminho destino para chegar a um determinado recurso.
- **Segurança** – Nem sempre vamos poder garantir a integridade do conteúdo que estamos a descarregar.

De modo a colmatar estas dificuldades, foi desenvolvida uma nova arquitetura que visa substituir a camada protocolar IP da Internet, denominada Information Centric Networks, na qual o foco deixa de ser «*Onde* arranjamos determinado conteúdo?», dando lugar à pergunta fundamental «*Qual* é o conteúdo do qual efetivamente estamos à procura?».

## II. NOMEAÇÃO DE CONTEÚDO

Atualmente a recuperação de conteúdo através da Internet segue um modelo intrinsecamente *host-centric*. Para estabelecer comunicação com o hospedeiro do conteúdo, um utilizador necessita de ter conhecimento prévio do seu endereço IP. Esta abordagem restringe o conteúdo a uma localização e identificação, sendo que a abordagem utilizada pela ICNs difere significativamente de este tipo de arquitetura.

As ICNs permitem o conteúdo ser requisitado através do nome, graças à utilização de esquemas de nomeação de conteúdos que deverão garantir a unicidade, persistência e escalabilidade, sendo eles:

- Nomeação Simples:

Os nomes simples, ou *Flat names*, são conjuntos de bits de tamanho fixo utilizados para identificar objetos, sendo que a abordagem mais comum no mapeamento destes identificadores é a utilização de uma função criptográfica de hash. Não existe qualquer tipo de relação entre o identificador do conteúdo e sua localização, proveniência

<sup>1</sup> Peer to peer

<sup>2</sup> Content delivery networks

ou qualquer outra característica além o vínculo entre o conteúdo em si e seu nome.

Este esquema garante a persistência graças à ausência de semântica e a unicidade graças à baixa probabilidade de colisão no mapeamento. No entanto, os nomes não são *user-friendly* e não garantem a escalabilidade, pois é impossível agregá-los por prefixo, o que leva a que as tabelas de encaminhamento necessitem de ter uma entrada por nome.

- Nomeação Hierárquica:

Os nomes hierárquicos são obtidos através da concatenação de nomes de componentes do conteúdo de modo a formar um identificador único. Os componentes refletem as propriedades do conteúdo (versão, formato, etc).

O acesso a este tipo de conteúdo dinâmico necessita que os utilizadores consigam construir os nomes do conteúdo desejado sem nenhum conhecimento prévio do seu nome ou do próprio conteúdo a obter. Pedidos relativos e nomes parciais podem ser utilizados para determinar as sequências de nomes, servindo-se das relações hierárquicas entre componentes. Ao contrário dos *Flat names*, os nomes hierárquicos podem ser agregados por prefixo mais longo, o que torna a transição entre o paradigma da Internet atual para o paradigma das ICNs mais fácil, visto que funciona de forma semelhante aos protocolos de encaminhamento IP. Apesar disso, este esquema não garante a persistência dos identificadores, pois mudanças efetuadas no conteúdo podem modificar os seus componentes e a alterar a sequência gerada.

- Nomeação Baseada em Atributos:

Os nomes baseados em atributos não atribuem um identificador único a cada conteúdo, mas sim um conjunto de pares atributo-valor (AVP). O conteúdo passa a ser então identificado pelos seus pares e o seu acesso é feito aplicando restrições aos mesmos. O conjunto de restrições que podem ser utilizadas para identificar conteúdos denominam-se predicados. Existe uma relação entre os predicados, o conjunto de restrições que os definem e os conteúdos que eles representam, chamado cobertura. Um predicado cobre um outro se todo o conteúdo obtido através último é possível obter através do primeiro, sendo esta propriedade o que permite a agregação dos AVPs. Visto que os conteúdos não são explicitamente nomeados, podemos especificar predicados que cobrem grandes conjuntos de conteúdo e delegar a responsabilidade de identificar o conteúdo pretendido ao utilizador que o requisita.

### III. ENCAMINHAMENTO BASEADO EM NOMES

Como mencionado anteriormente, as ICNs devem conseguir distribuir conteúdo requisitado através do seu nome, sem qualquer outro tipo de informação em relação à sua localização. Para tal se verificar, os nós da rede necessitam de informação acerca de todo o conteúdo existente na

mesma de modo a encaminharem os pedidos a cópias válidas desse conteúdo.

Este encaminhamento de conteúdo, ou encaminhamento baseado em nomes, tem características particulares no que trata a forma como a informação de encaminhamento é explorada pelos nós assim como a forma segundo a qual a mesma informação é guardada na rede. Estes mecanismos podem ser divididos em dois grupos principais: encaminhamento não-hierárquico e encaminhamento hierárquico.

- Encaminhamento Não-Hierárquico

No encaminhamento não-hierárquico, ou não estruturado, redes interconectadas são vistas como uma única rede, onde roteadores, *gateways* e *bridges* são apenas nós adicionais. Não é feito nenhum esforço de forma a organizar a rede e o seu tráfego. Ao estabelecer ligações entre os nós de acordo com as necessidades de entrega de conteúdo no momento, este tipo de encaminhamento permite que todos os nós obtenham conteúdo válido. Dado que a rede não está organizada de forma hierárquica, não existe um nó raiz ou um nó central encarregue de guardar informação de encaminhamento, fazendo com que essa informação seja difundida a todos os nós, permitindo assim que todos eles consigam calcular as melhores rotas para a entrega de conteúdo, pois cada nó tem informação sobre todos os outros nós da rede.

Este tipo de encaminhamento permite o uso de diferentes caminhos para o mesmo conteúdo, uma vez que a informação sobre a topologia de toda a rede permite o cálculo de rotas acíclicas e aumenta a disponibilidade da rede como um todo, dado que não pode existir um único ponto de falha. Os protocolos de encaminhamento utilizados na Internet são, no geral, não-hierárquicos. Como tal, a maior parte dos problemas encontrados neste tipo de protocolos já foram previamente identificados e trabalhados, tornando assim possível aplicar as soluções para os mesmos na utilização de encaminhamento não-hierárquico.

- Encaminhamento Hierárquico

O encaminhamento hierárquico permite a divisão de uma rede em domínios autónomo de encaminhamento. Um domínio de encaminhamento é uma parte da rede que é manipulada pela mesma instância de um protocolo de encaminhamento. Os roteadores que pertençam a um domínio não têm conhecimento sobre a topologia exata de outros domínios, apenas sabem a lista de destinos incluídos nesse domínio e os seus respetivos custos de acesso. Todos os roteadores de fronteira entre domínios têm visibilidade em cada um dos domínios que eles interconectam.

As ICNs possuem basicamente dois tipos de encaminhamento hierárquico: arquiteturas baseadas em árvores e arquiteturas de tabelas de hash distribuídas.

Topologias baseadas em árvores funcionam da seguinte forma:

Os nós Pai são aqueles que estão conectados a um ou mais nós filho, configurando assim a raiz da sub-árvore á qual o nó filho pertence. Os nós par, ou *peer nodes* são os nós que pertencem ao mesmo nível hierárquico em relação ao seu nó raiz comum.

Os nós par são diretamente acessíveis. Os nós pai possuem toda a informação de encaminhamento dos seus filhos que servem de *gateway* entre o pai e as sub-árvores dos filhos. É nos nós pai que fica agregada toda a carga de encaminhamento para a sub-árvore inteira, diminuindo a quantidade de informação utilizada por cada nó e, consequentemente, reduzindo os seus requisitos computacionais (energia, memória, etc). Os nós da rede não possuem um mapa completo da topologia da rede, como mencionado anteriormente. Eles apenas guardam informação relativa ao seu nó pai, nós pares e nós filhos.

Topologias de tabelas de hash distribuídas funcionam da seguinte forma:

Esta topologia permite o armazenamento e recuperação de informação associada a uma chave hash, numa rede de nós par. Os nós de uma rede cordenam-se entre eles para balancear e armazenar a informação existente na rede sem qualquer tipo de coordenação por parte de um nó central. Estes mesmo nós apenas guardam uma parte da informação de encaminhamento da rede, o que torna o processo de disseminação da informação mais simple quando um novo nó se junta á rede, visto que essa disseminação tem como destino um numero reduzido de nós.

Como cada nós apenas uma parte da tabela de encaminhamento, o processo de encontrar ou armazenar um par chave/valor requer o contacto com multiplos nós. Existem dois tipos diferentes mais comuns para operações de procura. Numa procura iterativa, o nó a fazer o pedido irá consultar um outro nó por um par chave/valor. Se esse nó não possui essa informação ele irá retornar um ou mais nós que estão mais “próximos”. O nó que fez o pedido irá então consultar o nó mais próximo, sendo que este processo se repete até o par chave/valor ser encontrado ou o último nó consultado retornar um erro, especificando que o par não pode ser encontrado. Numa procura iterativa, o nó que faz o pedido vai consultar o nó mais próximo, que por sua vez irá consultar o nó mais próximo, e assim sucessivamente.

Os custos de processamento e de cache envolvido no mapeamento das chaves são divididos entre todos os nós, garantindo uma proteção contra pontos únicos de falha, ao contrário das topologias baseadas em árvores, onde um único ponto de falha pode causar a perda de ramos inteiros da árvore. Mecanismos hierárquicos baseados em DHT, ou DHTs hierárquicas (H-DHT), permitem a organização de nós em redes de sobreposição de modo a encaminhar mensagens para as hash keys correspondentes de modo eficiente. As estruturas H-DHT garantem que todos os nós num domínio específico fazem parte de uma DHT exclusiva,

de modo a que níveis hierárquicos mais altos consistem na fusão dos níveis mais baixos.

#### IV. CACHE DE CONTEÚDO

Os roteadores de conteúdo podem ser estendidos de modo a providenciar uma infraestrutura de armazenamento distribuído, do mesmo modo segundo o qual CDNs tradicionais operam. Á medida que o conteúdo é redirecionado para diferentes nós, o router pode guardar em memória o conteúdo mais frequentemente acedido, operando assim como uma cache de rede. Apesar das similaridades, ao contrário de como funcionam as CDNs, nas ICNs a decisão do processo de cache é baseada apenas em informação local. Os nós têm em consideração pedidos de conteúdo e conteúdo entregue a quando da determinação do conteúdo a manter em cache, sendo que qualquer nó pode tomar funções de cache. Apesar de não lidarem diretamente com a localização do conteúdo, o uso de cache na rede acaba por distribuir copias de conteúdo para nós distantes, mais próximos dos utilizadores.

#### V. ICNs E A INTERNET DE HOJE EM DIA

#### VI. DESAFIOS PRÁTICOS DA IMPLEMENTAÇÃO DA ARQUITETURA ICN

Já discutimos as vantagens da arquitetura ICN nos dias atuais, nomeadamente o benefício que esta poderá trazer para o uso casual da Internet por um utilizador doméstico, o qual prioriza conteúdo de multimédia, e para aplicações que exijam alta escalabilidade de forma a servir um grande número de utilizadores concorrentes; vamos agora voltar o nosso foco para alguns desafios práticos advindos da implementação da arquitetura ICN.

O primeiro empenço que queremos denunciar é a adoção de modelos económicos que joguem com os requerimentos técnicos da execução desta arquitetura a grande escala. Contemporaneamente, as ISPs baseiam o seu modelo de negócio na conectividade, isto é, utilizadores subscrevem ao serviço de ISPs de forma a obterem acesso à Internet, as quais reenviam os pacotes que chegam até elas para outras ISPs. Como a Internet está organizada em redes autónomas, e os serviços da rede são *end-to-end*, também elas têm que pagar às outras ISPs pelo seu tráfego [2]. A eventual proliferação da arquitetura ICN na Internet implicaria uma mudança profunda no modo como as ISPs são remuneradas: estudos referidos em [5] descrevem um investimento inicial de capital bastante venerável, de forma a providenciar serviços de *caching*, que apesar do retorno financeiro posterior, não lhes proporcionam incentivo económico suficiente.

Voltemos agora a nossa atenção para o roteamento na Internet. No modelo tradicional, o envio de pacotes pelos roteadores é feito *end-to-end*, enquanto que numa ICN, o desenho destes aparelhos deve ser considerado com muito

mais cuidado, uma vez que deve implementar funções específicas da nova arquitetura, tais como a gestão de uma *cache* de conteúdo, que por sua vez inclui gestão do armazenamento e políticas de expiração do conteúdo, e *lookups* constantes por nós que pedem um determinado conteúdo [2]. É ainda de notar que no caso de uma *cache miss*, o conteúdo deve ser obtido de um nó auxiliar, que implica guardar informação de estado sobre pedidos pendentes no roteador de uma ICN.

Para concluir

#### BIBLIOGRAFIA

- [1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, e R. L. Braynard, «Networking named content», em *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, 2009, pp. 1–12.
- [2] I. M. M. Pedro B. Velloso Gabriel M. de Brito, *Information centric networks: a new paradigm for the Internet*, 1 ed. Wiley-ISTE, 2013.
- [3] «Image | Client-Server model». <https://upload.wikimedia.org/wikipedia/commons/thumb/c/c9/Client-server-model.svg/1200px-Client-server-model.svg.png>, Out-2019.
- [4] «RFC | Hypertext Transfer Protocol – HTTP/1.1». <https://tools.ietf.org/html/rfc2616>, Out-2019.
- [5] Y. Xu, Y. Li, S. Ci, T. Lin, e F. Chen, «Distributed caching via rewarding: an incentive caching model for ICN», em *GLOBECOM 2017-2017 IEEE Global Communications Conference*, 2017, pp. 1–6.